AD-A243 728

91-18889

198

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
Unclassified	N/A

2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT
N/A	Approved for Public Release
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	Distribution Unlimited
N/A	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
Utah State University	N/A	AFOSR/NL Building 410

6c. ADDRESS (City State and ZIP Code)	7b. ADDRESS (City, State and ZIP Code)
Logan, UT 84322-8200	Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
Same as block 7	N/A	Grant AFOSR – 89-0509

8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS.			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
	61102F	2312	A4	

11. TITLE (Include Security Classification) Environmental Containment Property Estimation Using QSARs in An Expert System

12. PERSONAL AUTHOR(S)
W.J. Doucette, M.S. Holt, D.J. Denne, J.E. McLean

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Annual	FROM 90/9/15 TO 91/10/15	1991/11/14	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Pollution, QSARs, Expert System, Organic Chemical, Property Estimation

19. ABSTRACT (Continue on reverse if necessary and identify by block number) A microcomputer program, utilizing molecular connectivity indices (MCI)-property, total molecular surface area (TSA)-property and property-property correlations and UNIFAC derived activity coefficients, is being developed to provide a fast, economical method to estimate aqueous solubility, octanol/water partition coefficients, vapor pressures, organic carbon, normalized soil sorption coefficients, bioconcentration factors, and Henry's Law constants for use in environmental fate modeling. The structural information for the MCI and UNIFAC models can be input using Simplified Molecular Input Line Entry System (SMILES) notation or connection tables generated from a commercially available two-dimensional drawing program. The TSA module accepts 3-D cartesian coordinates entered manually or directly reads coordinate files generated by molecular modeling software. In the MCI, TSA, and Property-Property modules, the user can select from either "universal" or class specific regression models for each property. To aid the user in choosing the most appropriate regression model(s), the program automatically suggests the most appropriate regression model based on the structure of the compound. In addition, the

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☑ SAME AS RPT ☐ DTIC USERS ☐	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
T. Jan Cerveny, Lt Col, USAF	(202) 767-5021	NL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

19

19. ABSTRACT (Cont'd)

statistics and list of compounds used in developing the model can be displayed. For the regression based modules, assessments of accuracy based on the 95% confidence interval and the estimated precision of the experimental values are provided along with the estimated property value. Additional correlation models can be easily added to PEP by the user. The database of measured properties, used in the development of the property estimation modules, and the Level 1 Fugacity Model are also linked directly to PEP. The current status and use of the program is described.

ENVIRONMENTAL CONTAINMENT PROPERTY ESTIMATION USING QSARs IN AN EXPERT SYSTEM

William J. Doucette
Mark S. Holt
Doug J. Denne
Joan E. McLean
Utah State University
Utah Water Research Laboratory
Logan, UT   84322-8200

October 15, 1991

Annual Report for Period 15 August 1990 - 15 October 1991

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

In order to assess the potential impact of the accidental introduction of an organic chemical into the environment, information is needed concerning its environmental fate. The fate of an organic chemical in the environment depends on a variety of physical, chemical and biological processes. Mathematical models, which attempt to integrate these processes, are widely used to predict the transport and distribution of organic contaminants in the environment. Use of these models requires a variety of input parameters which describe site and contaminant physical-chemical and biological characteristics. Several important contaminant properties used to assess the mobility and persistence of a chemical are aqueous solubility, octanol/water partition coefficient, soil/water sorption coefficient, Henry's Law constant, bioconcentration factor, and transformation rates for biodegradation, photolysis and hydrolysis.

One major limitation to the use of environmental fate models has been the lack of suitable values for many of these properties. The scarcity of data, due mainly to the difficulty and cost involved in experimental determination of such properties, has resulted in an increased reliance on the use of estimated values for many applications.

Quantitative Structure-Property Relationships (QSPRs) and Quantitative Property-Property Relationships (QPPRs) are methods by which properties of a chemical can be estimated from a knowledge of the structure of a molecule or from another more easily obtained property. Selection and application of the most appropriate QSPRs or QPPRs for a given compound is based on several factors including: the availability of required input, the methodology for calculating the necessary topological information, the appropriateness of a correlation to the chemical of interest, and an understanding of the mechanisms controlling the property being estimated.

A microcomputer program, utilizing molecular connectivity indices (MCI)-property, total molecular surface area (TSA)-property and property-property correlations and UNIFAC derived activity coefficients, was developed to provide a fast, economical method to estimate aqueous

1

solubility, octanol/water partition coefficients, vapor pressures, organic carbon normalized soil sorption coefficients, bioconcentration factors and Henry's Law constants for use in environmental fate modeling. The structural information for the MCI and UNIFAC modules can be input using Simplified Molecular Input Line Entry System (SMILES) notation or connection tables generated from a commercially available two-dimensional drawing program. The TSA module accepts 3-D cartesian coordinates entered manually or directly reads coordinate files generated by molecular modeling software. In the MCI, TSA and Property-property modules, the user can select from either "universal" or class specific regression models for each property. To aid the user in choosing the most appropriate regression model(s), the program automatically suggests the most appropriate regression model based on the structure of the compound. In addition, the statistics and list of compounds used in developing the model can be displayed. For the regression based modules, assessments of accuracy based on the 95% confidence interval and the estimated precision of the experimental values are provided along with the estimated property value. Additional correlation models can be easily added to PEP by the user. The database of measured properties, used in the development the property estimation modules, and the Level 1 Fugacity Model are also linked directly to PEP. The current status and use of the program will be described.

OBJECTIVES OR STATEMENT OF WORK

The primary goal of this project is to develop a microcomputer-based decision support system utilizing Quantitative Structure Property Relationships (QSPRs) and Quantitative Property Property Relationships (QPPRs) to predict the physical-chemical properties of an organic chemical which are necessary to model its environmental fate. The specific properties that are being investigated include: aqueous solubility (S), octanol/water partition coefficient (Kow), vapor pressure (Pv), organic carbon normalized soil/water partition coefficient (K_{oc}), Henry's Law constant (H), and bioconcentration factor (BCF).

In order to achieve the primary goal of this research, the following specific objectives are being accomplished:

1. Compile an accurate database of experimentally determined values of S, Kow, Pv, Koc, H, and BCF for a wide variety of organic compounds. Include compounds exhibiting a broad range of physical and chemical properties and expected mobility and persistence.

2. Using the database developed in Objective 1, evaluate and refine existing methods and/or develop new methods for estimating these contaminant properties using QSPRs and QPPRs.

3. Develop a microcomputer-based decision support system which incorporates the methods developed in Objective 2, to allow the prediction of environmental fate and transport properties of an organic contaminant upon inputting its structure. Provide an estimate of the accuracy of the predicted value using the decision support system.

4. Test the ability of the decision support system developed in Objective 3 to provide an accurate estimate of these environmental fate and transport properties. This will be done using a test set of chemicals for which accurate experimental values are available.

5. Compare the decision support system developed in Objective 3 to other widely used property estimation techniques.

BACKGROUND AND SIGNIFICANCE

Mathematical models are often used to estimate the fate and impact of organic chemicals in the environment. These models often idealize the environment as a system of connected compartments, i.e. water, soil, sediment, air and biota. The complexity of these models range from simple steady state models to non-steady state models which include a large number of compartments, transport between compartments and degradation processes.

Use of these models requires a variety of input parameters which describe site and contaminant physical-chemical and biological characteristics. Aqueous solubility (S), octanol/water partition coefficient (Kow), the organic carbon normalized soil/water sorption coefficient (Koc), vapor pressure (Pv), Henry's Law constant (H), and bioconcentration factor (BCF) are considered key properties used to assess the mobility and distribution of a chemical in environmental systems.

One major limitation to the use of environmental fate models has been the lack of suitable values for many of these properties. The scarcity of data, due mainly to the difficulty and cost involved in experimental determination of such properties for an ever increasing number of synthetic chemicals, has resulted in an increased reliance on the use of estimated values. Quantitative Property-Property Relationships (QPPRs) and Quantitative Structure-Property Relationships (QSPRs) have been used by environmental scientists and engineers to obtain estimated values for a variety of physical/chemical properties for use in environmental fate and assessment modeling.

Quantitative Property-Property Relationships (QPPRs), based on the relationship between two properties as determined by regression analysis, are used to predict the property of interest from another more easily obtained property. Frequently, the regression expressions are expressed in terms of the log of the two properties. Researchers have found that a number of environmental

properties can be related to one another in this manner. For example, QSPRs have been developed to estimate S, Koc and BCF from Kow and Koc and BCF from S [1-3].

Quantitative Structure-Property Relationships (QFPRs) are methods by which the properties of a chemical can be inferred or calculated from a knowledge of the structure of a molecule. QSPRs often take the form of a correlation between a structurally derived parameter(s) and the property of interest. For example, relationships between structurally derived parameters, such as molecular connectivity indices (MCIs) and total molecular surface area (TSA) and properties such as S, Kow, BCF, and H have been reported.

Molecular connectivity developed by Randic [4] and refined and expanded by Kier and Hall [5-7] is a method of bond counting from which topological indices, based on the structure of the compound, can be derived. For a given molecular structure, several types and order of MCIs can be calculated. Information on the molecular size, branching, cyclization, unsaturation and heteroatom content of a molecule is encoded in these various indices [5]. MCIs have been used to predict Koc [8,9], S [1], Kow [10], H [11] and BCFs [12].

A direct estimation of molecular surface area based on the concept of van der Waals radii, TSA, has been correlated with properties such as S, Kow, Pv and H [13-22]. Several different algorithms have been developed to calculate TSA which require the 3-D atomic coordinates of the solute molecule and the van der Waals radii of solute and solvent molecules as input [19,23].

Group contribution methods are another important category of QSPRs. The basic idea of a group contribution method is that while there is an enormous number of chemical compounds, both synthetic and naturally occurring, the number of functional groups that make up these compounds is much smaller. A single numerical value is assumed to represent the contribution of each functional group (i.e. a specified atom, a group of atoms bonded together or structural factor) to the physical property of interest. It is also usually assumed that the contributions made by each group are independent of each other. By summing up the values of the various fragments or groups the property of interest can be directly calculated.

The UNIFAC (UNIQUAC Functional Group Activity Coefficient) group contribution method [24-26], originally developed to estimate liquid phase activity coefficients in mixtures of nonelectrolytes, has been used by environmental researchers to estimate S and Kow [27-31]. In this technique, the activity coefficient is divided into two parts, a combinatorial part which reflects the size and shape of the molecule present and a residual portion which depends on functional group interactions. Various parameters, such as van der Waals group volumes and surface areas and group interaction parameters, are input into a series of equations from which the combinatorial and residual parts are calculated. Values for the group parameters have been tabulated and can be found in the literature[25,26]. UNIFAC is specifically designed to take into account interactions between groups and is appropriate for multiple solute/solvents systems. UNIFAC also permits estimates to be made as a function of temperature.

In most cases, more than one estimation method is available for a particular property. Estimation methods however, have widely varying accuracies and indiscriminate use of these techniques can result in large errors. Selection and application of QSPR or QPPR methods requires varying degrees of expertise that depend on the structure of a particular chemical of interest, knowledge of the mechanism of the process, the extent of the database used to develop the QSPR or QPPR and the complexity of the structural analysis required to relate structure to the property. For example, some QSPR and QPPRs are broader than others in the range of chemicals that are covered, and some methods have been established with a better understanding of the mechanisms or properties involved. In many cases estimation methods are developed from empirical or semi-empirical correlations. The success of the correlation is dependent on many factors including the type and number of compounds used in its development.

Incorporation of QSPR and QPPRs into a computer format is a logical and necessary step to gain full advantage of the methodologies for simplifying fate assessment. A practical computerized property estimation program, utilizing QSPR and QPPRs, should include the following attributes: be simple and flexible to use for both experts and non-experts, include

6

sufficient statistical information regarding the development of the QSPRs and QPPRs so that the range of applicability of such models can be evaluated and provide an indication of the accuracy of the estimated property.

A microcomputer based Property Estimation Program (PEP), utilizing MCI-property, TSA-property and property-property correlations and UNIFAC derived activity coefficients, is being developed to provide both experts and non-experts with a fast, economical method to estimate a compound's S, Kow, Pv, Koc, BCF, and H for use in environmental fate modeling. The user can input the required structural information for the MCI and UNIFAC calculation routines using either Simplified Molecular Input Line Entry System (SMILES) notation or connection tables generated ChemDraw™ a commercially available two-dimensional drawing program.. The TSA module accepts 3-D atomic coordinates entered manually or directly reads coordinate files generated by molecular modeling software such as Alchemy II™ or Chem3D Plus™. For property-property, TSA-property and MCI-property modules, the user can select from either "universal" or class specific regression models. To aid the user in choosing the most suitable regression model, the program automatically suggests the most appropriate regression model(s) based on the structure of the compound. In addition, the statistics associated with each model can be displayed along with the list of compounds used in developing the model. For the regression based modules, assessments of accuracy based on the 95% confidence interval and estimated precision of the experimental values are provided along with the estimated property value. Additional correlation models can be easily added to PEP by the user.

A chemical property database (PEP.DB), containing experimental values of S, Kow, H, Pv, Koc, and BCF complied from a variety of literature sources and computerized databases was used for developing the MCI-property, TSA-property and property-property relationships used in PEP. This database, which currently contains over 700 chemicals, is linked directly to PEP and provides the means for the user to search for chemical compounds by full or partial name or

synonym, to sort the compounds by name, boiling point, melting point, or molecular weight, and the ability to transfer to any of the property estimation modules.

A prototype database, containing information regarding the biodegradability of organic compounds is also being developed for incorporation into the PEP software system. This database, currently containing information for 33 chemicals, will be used to develop and evaluate relationships between structure and biodegradability. If successful, the resulting structure-biodegradability relationships will be incorporated into PEP during the third year of the project.

To illustrate the potential application of PEP the property estimation modules are linked directly to the Level 1 Fugacity Model developed by Mackay [32]. This simple model calculates the equilibrium distribution of an organic chemical between water, air, soil, sediment, suspended sediment and biota phases in a user defined world. The combination of PEP and the Fugacity Level 1 model provides the user with a methodology for predicting the environmental distribution of an organic chemical in a multi-phase system requiring only the structure of the chemical of interest as input. The current status and use of the PEP system will be described.

Description of Property Estimation Program (PEP)

HyperCard

The PEP software system is a HyperCard™ based program that runs on Apple Macintosh computers. HyperCard, which is bundled with every Macintosh sold, offers graphics, information storage, the means to display information in a variety of formats, the ability to establish links between related information, a high level language (HyperTalk), the ability to extend HyperTalk by writing new commands in a compiled language (i.e. C or Fortran) and a mechanism to transfer control to other Macintosh applications. The PEP system uses all these features.

HyperCard treats each screen full of information as a card and each set of related cards as a stack. Cards can contain fields for data and buttons for action procedures to operate on the data in the fields. This allows the standard Macintosh interface to be used without the direct use of the Macintosh toolbox routines greatly simplifying programming. In order to create a user interface the programmer simply draws, or creates the buttons or fields that are to be used. The link between buttons, fields and cards is done through HyperTalk. HyperTalk is an object oriented, interpreted language which allows the programmer to direct the flow of the program and at the same time allows the user the freedom to use the program as desired. However, large repetitive tasks and complicated computations can be very slow if HyperTalk is used. HyperCard also allows the programmer to link external functions or commands which are written in a conventional programming language as a means to speed up the slow interpreted language and implement custom features.

PEP requires the following system configuration to run: a Macintosh Plus, Macintosh SE, or Macintosh II computer, with a hard disk; HyperCard 2.0 software; Macintosh system software

version 6.0.5 or greater, running under MultiFinder; and a minimum of 2 megabytes of memory (RAM), with 1000 kBytes of memory allocated for HyperCard.

The PEP system currently consists of five HyperCard stacks: PEP Processor, PEP Models, PEP Help, Chemical Property Database and Biodegradation Database. While differing in purpose and characteristics, each stack uses a consistent set of icons and underlying programming technique. Table I lists some icons and button types and their uses.

Table I Symbols and button types used in PEP

Icon	Title	Action
	Return Arrow	Takes you back to the card you were at prior to this one
	Help	Shows Help for the current card
	Information	Shows general information
	Eye	Shows the equations or statistics
	Book	Shows the reference
	First Card	Takes you to the opening card of HyperCard
Pop Up Button		Lets the user choose from a popup menu list
Action Button		Initiates some acttion or calculation
⊠ **Check Box**		Select one or more from a list
◉ **Radio Button**		Select only one from a list

PEP Processor

This stack, which is divided up into four sections or modules, contains the algorithms for data input, calculations and output of the estimated physical-chemical properties. The opening screen of PEP is in the form of a flow chart, allowing the user to see the different modules and the overall organization of the program. Similarly, each module described in the following sections, is also organized in a flow chart form.

MCI Module Upon entering the MCI module, illustrated in Figure 1, the user must first input the necessary structural information using either SMILES notation [33,34] or connection tables generated from ChemDraw™, a commercially available, Macintosh compatible two-dimensional (2D) drawing program. SMILES is a chemical notation language specifically designed for computer use. It is a method of "unfolding" a 2D chemical structure into a single line of characters containing the structural information.

Figure 1 The MCI method card from PEP

After the structural information is entered, MCIs can then be calculated using a set of HyperCard™ external functions (XFCN) written in the programming language C based on code described by Frazier [35]. The MCI calculation routine in PEP calculates simple, bond and valence indices of several types (path, cluster, chain, and path/cluster) and orders (0 through 6), if possible, for each molecule. This results in a maximum of 54 index values for each molecule which can be displayed on screen and/or output to a printer. To account for non-dispersive force effects on aqueous solubility and solubility related properties, zero through six order Δ valence path indices ($\Delta\chi$), as described by Bahnick and Doucette [36], are calculated by PEP, in addition to the 54 indices described above. To calculate $\Delta\chi$ indices, a nonpolar equivalent is made by

12

substituting C for O or N atoms. MCIs are calculated for the nonpolar equivalent and values for $\Delta\chi$ can be computed for each type of index by:

$$\Delta\chi = (\chi)np - \chi \qquad (1)$$

After the MCIs are calculated, they can be displayed or printed if desired and the user can then choose which properties are to be estimated. For each property, two categories of MCI-property relationships, "PEP" and "LIT", are displayed. The PEP category provides a list of all MCI-property relationships, both class specific and "universal", that were developed in this project using the experimental values reported in the PEP property database. "Universal" MCI property relationships were developed using all available experimental data for a given property regardless of chemical class. Class specific MCI-property relationships were developed if property values were available for a sufficient number (10 or greater) of compounds within a particular chemical class (PCBs, PAHs, ureas, etc.). In addition, several multi-class MCI-property correlations were developed for more broad classes of compounds such as halogenated aliphatics, halogenated aromatics, etc. To aid the user in choosing the most appropriate model, the suggested chemical classes based on structure are flagged with a diamond in the popup menu. The chemical class is determined by looking for a group of atoms and bonds between the atoms that distinguish a chemical class. The number of chemical classes that are chosen by the program will be the number of different distinguishing subgroups that are found. In addition, a summary of the regression statistics and list of compounds used to develop and evaluate each MCI-property relationship can be displayed by clicking the "eye" or "view statistics option" found at the left of each regression model. An example of the statistical information provided for each MCI-property relationship is shown in Figure 2. Information displayed on the statistics card includes: the MCI-property regression equation, the list of chemicals used in developing the regression model, the standard errors of the coefficients in the regression equation, the Analysis of Variance (ANOVA) table, the

13

r2 value, a graph of the predicted vs. estimated values, a graph of the residuals vs. the predicted values, a graph of the residuals vs. the number of standard deviations and appropriate reference.

| **   File   Edit   Go   Print   Misc.** | | | | | | | **?**   ⏎ |

STATISTICS Class: S Universal

Regression Results						Analysis of Variance Table				
Variable	**Coef.**	**Std. Error**	**t**			**Source**	**RSS**	**d$_f$**	**MSS**	**F**
Constant	0.3917	0.1376	2.85	⬆		Regression	889.176	2	445	446
vp1	-.9257	0.0316	-29.3			Residual	360.920	362	0.997	
Δvp1	1.8251	0.1047	17.4	⬇		Total	1250.096	364	3.4343	

$r^2 = 71.1\%$ $n_{obs} = 365$ $S = 0.9985$

Predicted vs. Exp.   **Residual vs. Predicted**   **Residual vs. Prob.**

Figure 2 Example statistics card from PEP

The "LIT" MCI-property correlations were complied from various literature sources. Information regarding the number and type of compounds included in these models is provided if it was available in the original reference. Clicking on the "book" icon will display the reference of the "LIT" regressions.

After choosing the most appropriate regression, estimates for the selected properties can be made. As shown in Figure 3, the MCI module results window provides an estimate of the property along with its calculated accuracy based on both the 95% confidence interval calculated

14

from the regression and the estimated precision associated with the experimental determination of the property. In addition, the user has the option to search the property database for actual experimental values if they are available for comparison.

⬥ File Edit Go Print Misc. Models					[?] ⏎

Estimated Properties

Chemical Name: chlorobenzene

Method: MCI

Property	*Value	Units	Regression Equation Used	View stats	Look for values in Prop.Db
log antilog					
log S	-1.90 ± 1.96	moles/L	PEP: Universal	👁	-2.360 moles\Liter
log Kow	2.95 ± 1.43		PEP: Universal	👁	2.840
log Pv	2.86 ± 1.19	Pascals	PEP: Halogenated Aromatics	👁	1.075 mm Hg
log H	-.84 ± 1.59	dimensionless	PEP: Universal	👁	-0.818 dimensionless
log Koc	2.31 ± 0.67		PEP: Halogenated Aromatics	👁	2.410
log BCF	2.45 ± 1.41		PEP: Universal	👁	NA

*Note: The values shown are estimated at 25°C ± the 95% prediction interval for the regression used.

Figure 3 Results card from PEP

TSA module The TSA module is similar in operation to the MCI module. The user must first input the required structural information. However, unlike MCIs, the calculation of TSA requires information describing the geometry of the molecule in terms of its 3-D atomic coordinates. The TSA module accepts 3-D atomic coordinates entered manually or directly reads coordinate files generated by commercially available, Macintosh compatible, molecular modeling software such as Alchemy II™ or Chem3D Plus™. The TSA module is also designed to accept files generated by other hardware/software combinations including CONCORD (Tripos

Associates, Inc.), a hybrid expert system and molecular modeling software designed for the rapid generation of high quality approximate 3-D molecular structures. In addition to the 3-D molecular structure, the user must also input van der Waals radii for each of the atoms. A editable table of van der Waal radii, obtained from Bondi [37] for most common atoms, is provided within the TSA module. Once the molecular geometry and the van der Waal radii are input, TSA can be calculated using a XFCN which was adapted from the SAVOL2 algorithm developed by Pearlman [19]. In this algorithm, each atom of a molecule is represented by a sphere centered at the equilibrium position of the nucleus. The radius of the sphere is equal to that of the van der Waals radius. Planes of intersection between spheres are used to estimate the contribution to surface area from the individual atoms or groups. The program computes the surface area of individual atoms or group by numerical integration, and the overlap due to intersecting spheres is excluded from the calculation. TSA is calculated by the summation of individual group contributions. The program also allows the TSA of the solute molecule to be calculated after the addition of a suitable solvent radius. A more detailed description of the TSA calculation method is provided by Pearlman [19].

After the TSA has been calculated, the user then chooses the properties of interest and a regression equation for each using the same approach as described in the MCI module. If the SMILES string or the connection table is also input, the appropriate chemical classes will be flagged in the popup menu. The operation of the TSA module from this point on is identical to that of the MCI module.

UNIFAC module Like the MCI module, the UNIFAC module (illustrated in Figure 4) requires either a SMILES string or a connection table as input. An XFCN converts the structural information provided by the SMILES string or connection table into valid UNIFAC subgroups and counts the number of each subgroups present. In order to break the structure into the proper subgroups, the SMILES string or the connection table is interpreted and the information is put into a matrix. Each row and column in the matrix represents an atom in the chemical. The matrix

16

contains the bond order between the two atoms that corresponds to each entry in the matrix. If two atoms are not connected then a 0 is placed in the corresponding entry in the matrix. After the matrix is built the algorithm then "asks" specific questions about each atom, its neighbors, and how it is connected. If the answers to a set of questions are all true then a subgroup was found, the atoms are put together and the matrix is reduced. The questions are then repeated and the next subgroup is chosen, this continues until no additional subgroups are found. The questions are asked in a specific sequence so that the resulting subgroups are independent of the order of the atoms in the matrix.

Figure 4 UNIFAC method card from PEP

The UNIFAC method for calculating activity coefficients, as described by Grain [38], is implemented using both HyperTalk and an XFCN. The functional group interaction parameters,

17

presented by Gmehling et al. [26] and derived from vapor-liquid equilibria (VLE), are used in the calculation routine but can be changed by the user. After the activity coefficients are calculated they can be displayed along with relevant intermediate values and used to estimate S and Kow by the following expressions:

$$Kow = 0.115 \, \gamma_{\infty w} / \gamma_{\infty o} \tag{2}$$

$$S \, (mol/L) = 55.6 / \gamma_{\infty w} \tag{3}$$

where $\gamma_{\infty w}$ is the activity coefficient of the chemical infinitely dilute in water and $\gamma_{\infty o}$ is the activity coefficient of the chemical infinitely dilute in octanol [27].

Property/Property Module   Input for the Property/Property module, shown in Figure 5, depends on the properties to be estimated and the regression models used. Thus, the user must select the properties to be estimated and the regression equations to be used before any input values are requested. The program keeps track of the inputs required and provides the appropriate input fields. If available, the required properties can be imported directly from the associated chemical property database. Information regarding the regression statistics, if available, is also provided as previously described in the MCI module. After the necessary properties are entered into the corresponding input fields, the properties of interest can be estimated and the results, along with the 95% prediction interval (if the necessary data is available) can be viewed.

18

Property/Property

Chemical Name: chlorobenzene

1. Choose Property	2. Choose Regression	3. Input Properties

1. Choose Property

☒ S

☒ Kow

☒ Koc

☒ BCF

2. Choose Regression

| PEP | |
| LIT. | Hansch et al. 1968 from Kow |

| PEP | Universal from S |
| LIT. | |

| PEP | |
| LIT. | Karickhoff et al. 1979 from Kow |

| PEP | Universal from Kow |
| LIT. | |

3. Input Properties

Look in Prop. DB

log S -2.360   moles\L

log Kow 2.840

log P

log H

log Koc

log BCF

M. P.

M.W.

4. Estimate Properties

Note: All values are at 25°C

Figure 5 Property/Property correlation method from PEP

PEP Models

To illustrate the practical application of PEP, an additional stack called PEP Models was developed. This stack, which contains the algorithm for the Level 1 Fugacity Model [32], is linked directly to the PEP Processor, but can also be used independently. The Level 1 Fugacity Model considers a unit world consisting of six compartments: air, water, soil, suspended solids, sediment, and biota. The model predicts the equilibrium concentrations of the chemical of interest in each compartment using the fugacity approach described by Mackay [32]. The model requires the input of Koc, H and BCF which can be read directly from the PEP processor or the PEP

chemical property database, if available. In addition to the chemical specific properties, the density and volume of each compartment must be specified along with the organic carbon content of the soil, sediment and suspended sediment. An editable set of default values for compartment density, volume and organic carbon content, as suggested by Mackay, is provided. A complete description of the model has been given by Mackay [32,39].

File Edit Go Print	↵ ⑦

Fugacity Level 1

Chemical Name: chlorobenzene

1. Property Values	2. Input Environmental Compartment Values
Look for Values in Prep.DB	Save Current Values Values: [Unit World.Default]
	Delete Saved Values [Set Default Values]

	Compartment	Density kg/m³	Volume m³	% Organic Carbon
log Koc ⊠ **2.410**	⊠ Air	1.19	1e10	
log H ⊠ **(dimensionless)** **-0.818**	⊠ Water	1000	7e6	
log BCF ☐ **2.45**	⊠ Soil	1500	9e3	2
	⊠ Susp. Solids	1500	35	4
⊠ Indicates value from DB	⊠ Sediment	1500	2.1e4	4
	⊠ Biota	1000	3.5	

3. Calculate Distribution

Figure 6 PEP Models card

After the user inputs all the required information, the model calculations are performed in HyperTalk and the results are presented in both in tabular and graphical form as illustrated in Figure 7. The results can be viewed graphically in either bar or pie chart forms in terms of the concentration or percent of the chemical in each phase.

Fugacity Results

Chemical Name: chlorobenzene

	Percent in each compartment	Conc. mol/m³
Air	99.51	9.95E-9
Water	0.458	6.54E-8
Soil	0.004	5.05E-7
Susp. Solids	0.000	1.01E-6
Sediment	0.021	1.01E-6
Biota	0.000	1.84E-5
Total	100.0	

Values used

Koc	2.410	☒
H	-0.818	☒
BCF	2.45	☐

☒ indicates values from DB

Graph: Concentration

Concentration in each Compartment

Air
Water
Soil
Susp. Solids
Sediment
Biota

0   10   20
moles/L ×10⁶

Figure 7 PEP Models results card

PEP Help

Information regarding the operation of each property estimation module and the chemical property database are available in the PEP Help stack which is easily accessed at any time within the PEP system. The organization and layout of each help card is similar to that illustrated in Figure 8 for the MCI module. The user can select the topic of interest by clicking on the appropriate radio button and the information on that subject will be displayed in the scrolling field.

MCI Module

MCI Options

- ◉ overview
- ○ input
- ○ calculating MCIs
- ○ choosing the model & the chemical class
- ○ property estimation
- ○ limitations

> Molecular connectivity developed by Randic (1972) and refined and expanded by (Kier and Hall, 1976, 1980, 1986) is a method of bond counting from which topological indexes, based on the structure of the compound, can be derived. For a given molecular structure, several types and orders of molecular connectivity indexes (MCIs) can be calculated. Information on the molecular size, branching, cyclization, unsaturation, and heteroatom content of a molecule is encoded in these various indices (Kier and Hall, 1976). Molecular connectivity has been used to predict Koc (Sabljic, 1984, Sabljic, 1987, Bahnick and Doucette, 1988), S (Doucette, 1985, Nirmalakhandan and Speece, 198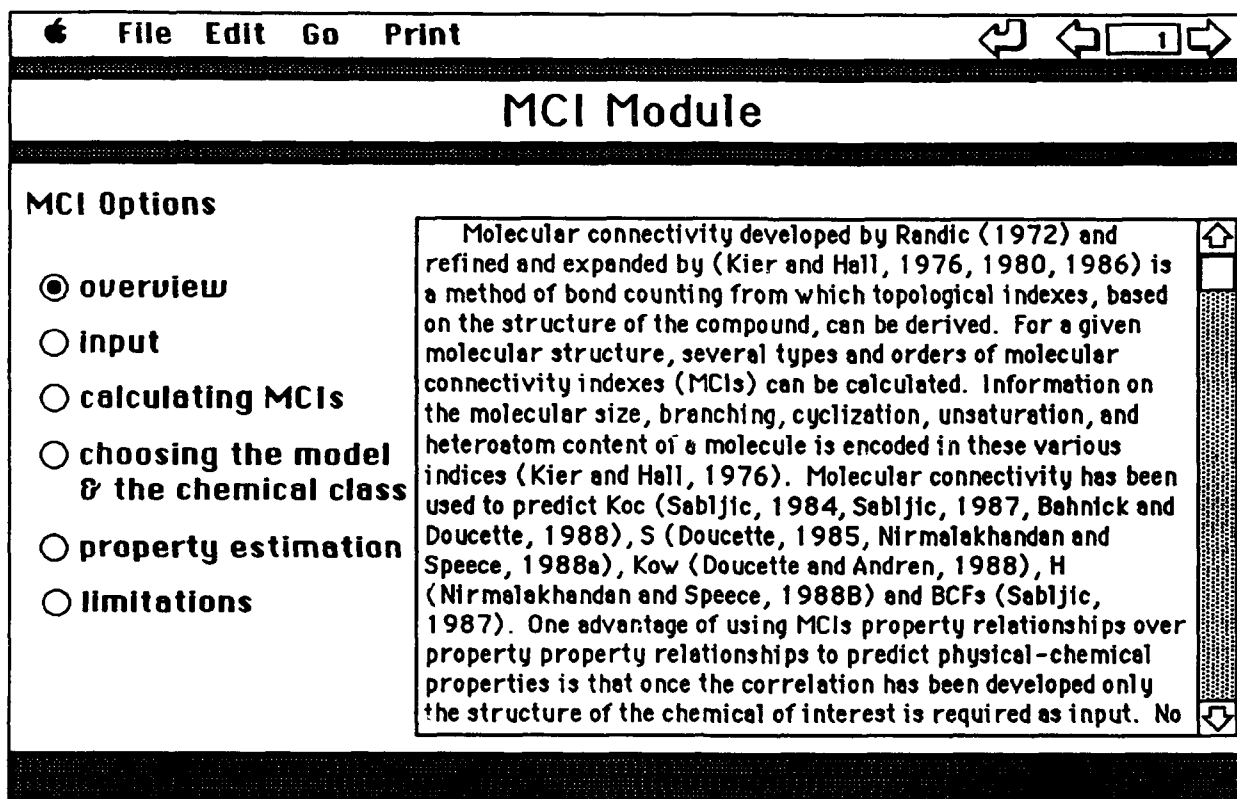8a), Kow (Doucette and Andren, 1988), H (Nirmalakhandan and Speece, 1988B) and BCFs (Sabljic, 1987). One advantage of using MCIs property relationships over property property relationships to predict physical-chemical properties is that once the correlation has been developed only the structure of the chemical of interest is required as input. No

Figure 8 Example Help card from PEP

Chemical Property Database

Experimentally determined physical property data for about 700 compounds, having at least one value of aqueous solubility (S), octanol/water partition coefficient (Kow), vapor pressure (Pv), organic carbon normalized soil sorption coefficient (Koc), bioconcentration factor (BCF), or Henry's law constant (H), was complied from a variety of literature sources and computerized databases. Using this information, a chemical property database was constructed using HyperCard™ and subsequently used for developing MCI-property, TSA-property and property-property relationships. In addition to the properties listed above, the database includes the

following information: compound name and synonyms, a diagram of the 2-D chemical structure, CAS number, chemical formula, molecular weight (MW), boiling point (BP), melting point (MP), and appropriate references for each value. A built-in unit conversion utility enables users to quickly view property values in a variety of commonly used units. The database is directly connected to the PEP Processor stack.

The Chemical Property Database also provides the means for the user to search for chemical compounds by full or partial name or synonym, to sort the compounds by name, boiling point, melting point, or molecular weight, and the ability to transfer to any of the property estimation modules. In addition, the user can easily edit exiting values or add new information. The chemical property database screen is illustrated in Figure 9.
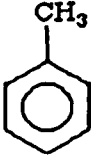

Figure 9 Example card from the Chemical Property Data Base

Biodegradation Database

A prototype database, containing information regarding the biodegradability of organic compounds is also being developed for incorporation into the PEP software system. This database, currently containing information for 33 chemicals, will be used to develop and evaluate relationships between structure and biodegradability. If successful, the resulting structure-biodegradability relationships will be incorporated into PEP during the third year of the project. In its current state, the biodegradation database contains the following information: compound name, structure, SMILES string, molecular weight, aqueous solubility, octanol/water partition coefficient, reference, matrix (soil, culture), study type (microcosm, field, liquid culture), endpoint (mineralization, disappearance of parent compound, identification of intermediate degradation product), chemical concentration, percent loss of chemical over time, degradation rate constant and order, half-life, organism(s), and environmental variables (temperature, pH,% soil organic matter).
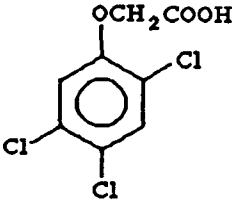
2,4,5-trichlorophenoxyacetic acid

CAS Number: 93-76-5
Formula: C8H5Cl3O3

FULL REFERENCE

Gibson, S. A. and Suflita, J. M. 1990. Anaerobic biodegradation of 2,4,5-trichlorophenoxyacetic acid in samples from a methanogenic aquifer : stimulation by short-chain organic acids and alcohols. Applied and Environmental Microbiology. 5(6) : 1825-1832.

structure

OCH_2COOH

Cl
Cl
Cl

MATRIX	STUDY TYPE	O$_2$ CONDITIONS	ENDPOINT
soil	soil microcosm	anaerobic	transformation

CHEMICAL CONC.	% LOSS/TIME	RATE CONSTANT (k)	ORDER	t 1/2
300-500µM	85/12 weeks			

ORGANISM(S)

Anaerobic sediment from anoxic aquifer

COMMENTS

Autoclaved controls showed no disappearance of 2,4,5-T nor the appearance of degradation products. After twelve weeks of incubation, about 85% of the parent material could be accounted for as a series of

ENVIRONMENTAL VARIABLES

temperature (°C): ?
CEC: ?
pH: ?
%OM: ?

| find | sort | add | delete | print | export | PEP | PEP db | ? | | | | 4 | |

Figure 10 Example Card from PEP Biodegradation Database

25

SUMMARY

A microcomputer program for estimating physical/chemical properties of organic chemicals for use in environmental fate modeling has been described. The Property Estimation Program (PEP) and associated physical property database (PEP.DB) was developed using HyperCard for the Apple Macintosh series of computers. The PEP system utilizes both quantitative structure-property and property-property relationships (QSPRs and QPPRs) to provide the user with several approaches to estimate S, Kow, Pv, H, Koc and BCF depending on the information available. While QPPRs have been used by both experts and non-experts for estimating properties, one of the major limitations in using QSPRs has been the difficulty in using the necessary software tools. The graphical interface and flow chart design of PEP leads the user through a series of logical steps designed to provide even non-experts with a economical, easy to use software system for property estimation. The structural information for the MCI and UNIFAC modules can be input using Simplified Molecular Input Line Entry System (SMILES) notation or connection tables generated from a commercially available two-dimensional drawing program. The TSA module accepts 3-D cartesian coordinates entered manually or directly reads coordinate files generated by molecular modeling software. For each property the user can select from either "universal" or class specific regression models. The program's built in intelligence helps the user choose the most appropriate QSPR based on the structure of the chemical of interest. In addition, sufficient statistical information is provided to allow the user to determine on the validity of the QSPRs and QPPRs utilized in PEP. Designed to make the program both practical and educational, on line documentation is provided not only for the operational characteristics of the program but also for the theory associated with the property estimation techniques. The combination of the various property estimation methods, chemical property database, and simple environmental fate model (Level 1 Fugacity Model) illustrates the potential application of PEP in both educational and regulatory settings.

SUMMARY OF SECOND YEAR ACCOMPLISHMENTS
AND THIRD YEAR OBJECTIVES

PEP improvements and modifications

1. Chemical Property Database

a) Database layout and user interface improved. Pop up buttons for database functions were replaced with pull down menus to make PEP more like standard Macintosh applications.

b) Added additional information for approximately 100 new compounds.

c). Added export feature that allows user to export all information in database as text.

d) Added report feature that allows user to export all information on specific card (i.e. specific chemical) as text.

e) Added feature that enables user to ability to add and delete references from database.

In summary, the PEP chemical property database is fully functional. No additional features or changes to user interface are planned for the final year of the project. New chemical property data will be added as it becomes available.

2.PEP Processor

a) MCI module

1) Developed decision support system for choosing most appropriate QSPR based on chemical class.

2) The MCI calculation algorithm was changed from an C application external to Hypercard to a Hypercard external function. This change yielded an approximate ten fold increase in the efficiency of the MCI calculation algorithm. Versions were complied for machines with and without a Floating Point Processor (FPU).

3) MCI module modified to accept both hydrogen suppressed and non-hydrogen suppressed connection tables automatically.

4) User interface changed to flow chart format for greater ease of use and consistency between modules.

In summary, MCI module is fully functional. In the final year of the project, we will continue to refine the MCI-property relationships. The relationship between MCIs and properties such as polarizability, dipole moment, partial atomic charge, and linear solvation parameters will be investigated.

b) UNIFAC module

1. Completed decision support system for dissecting SMILES strings or connection tables into appropriate UNIFAC groups and retrieving necessary group values from UNIFAC database.

2. User interface changed to flow chart format for greater ease of use and consistency between modules.

In summary, the UNIFAC module is fully functional. No changes to the user interface are planned for the final year of the project. During the third year of the project, the validity of the UNIFAC estimates will be examined using a test set of compounds having experimentally measured physical property data. Property estimates from the UNIFAC model will also be compared to estimates made with the other PEP modules and with other literature methods. One additional property, the oil/water partition coefficient, will also be investigated and incorporated into the UNIFAC module.

c) Property property module

1. User interface changed to flow chart format for greater ease of use and consistency between modules. In the final year of the project, additional property-property

28

correlations, obtained in this study and from the literature, will be implemented. In addition, the validity of the Property-property relationships will be examined using a test set of compounds having experimentally measured physical property data.

c) TSA module

1) Converted VAX fortran version of Pearlman's SAVOL2 [19] algorithm for calculating TSA to C application on Macintosh, then converted C application to Hypercard external function.

2) Investigated both universal and class specific TSA-property relationships and implemented a prototype version of TSA module that uses TSA-property relationships to estimate S, Kow, Koc, Pv, H, BCF.

3) Investigated the relationship between TSA and partial atomic charge and S, Kow, Koc, Pv, H, BCF .

During the third year of the project, the TSA-property relationships will be refined and finalized and the validity of the resulting relationships will be examined using a test set of compounds having experimentally measured physical property data.

d) PEP Help

1) Added help module that users can access from any location in the PEP software system. The PEP Help module provides the user with information regarding the operation of the various PEP modules. In addition, the PEP Help module also provides information regarding the calculation of MCIs, TSA, and UNIFAC-derived activity coefficients and the subsequent development of associated QSPRs.

e) PEP models

    1) Implemented Level 1 Fugacity Model developed by Mackay [3] into Hypercard and linked it directly to the property estimation modules.

f) PEP biodegradation database

    1) A prototype database, containing information regarding the biodegradability of organic compounds, is being developed for incorporation into the PEP software system. This database, currently containing information for 33 chemicals, will be used to develop and evaluate relationships between structure as described by MCIs and TSA and biodegradability. If successful, the resulting structure-biodegradability relationships will be incorporated into PEP during the third year of the project.

MISCELLANEOUS PUBLICATIONS

Frazier, J.D. 1990. Estimation of Chemical/Physical Properties Using Molecular Connectivity Indices for Application in Modeling the Environmental Fate of Organic Compounds. M.S., Utah State University.

Holt, M.S. and W.J. Doucette. 1991. Microcomputer program utilizing Quantitative Structure Property Relationships (QSPRs) and Quantitative Property Property Relationships (QSPRs) for the Estimation of physical/chemical properties used in environmental fate modeling In Proceedings of the 9th National Conference on Microcomputers in Civil Engineering, October 31, Orlando, Florida

Denne, D.J. (expected Dec. 1991). Investigating the Relation Between Molecular Structure and Parameters Important in the Environmental Fate of Organic Chemicals. M.S., Utah State University.

Holt, M.S. (expected Dec. 1991). Microcomputer Program Utilizing Quantitative Structure Activity Relationships for the Estimation of Properties for the Use in Fate and Transport Models. M.S., Utah State University.

LIST OF PAPERS/POSTERS PRESENTED AT PROFESSIONAL MEETINGS

Doucette, W.J., D.J. Denne, M.S. Holt and J.E. McLean. 1991. Prediction of Physical/Chemical Properties Using MCI and TSA-Property Relationships. Presented in platform session at the 12th Annual SETAC meeting, November 3-7, Seattle, WA..

Holt, M.S. and W.J. Doucette. 1991. Microcomputer program utilizing Quantitative Structure Property Relationships (QSPRs) and Quantitative Property Property Relationships (QSPRs) for the Estimation of physical/chemical properties used in environmental fate modeling Presented in platform session at the 9th National Conference on Microcomputers in Civil Engineering, October 31, Orlando, Florida

Doucette, W. J., D.J. Denne, J.D. Frazier, R.R. Dupont, D.K. Stevens, M. E Holt, and J.E. McLean. 1990. Microcomputer program utilizing quantitative structure-property and property-property relationships for environmental fate modeling. Presented in poster session at 11th Annual SETAC meeting, November 11-15, Arlington, VA, .

Doucette, W.J., D.J. Denne, J.D. Frazier, R.R. Dupont, D.K. Stevens, M. E Holt, and J.E. McLean. 1990. Microcomputer program utilizing molecular connectivity indices, UNIFAC-derived activity coefficients, and property-property correlations for the estimation of physical properties. Presented in poster session at the 4th International Workshop on QSARs in Environmental Toxicology, September 16-20, Veldhoven, The Netherlands

Doucette, W.J., H. Fugate, J.D. Frazier, and D.A. Bahnick. 1989. Structure Property Relationships Utilizing Molecular Connectivity Indices and Total Molecular Surface Areas for Environmental Fate Modeling. Presented in platform session at the International Chemical Congress of Pacific Basin Societies, December 21, Honolulu, Hawaii, USA

31

Frazier, J.D, and W.J. Doucette. 1989. Microcomputer Program Utilizing Molecular Connectivity Indices for Property Estimation. Presented in platform session at the International Chemical Congress of Pacific Basin Societies, December 21, Honolulu, Hawaii, USA

LIST OF GRADUATE STUDENTS ASSOCIATED WITH THE RESEARCH EFFORT

Joe Frazier, M.S. 1990 Thesis title: Estimation of Chemical/Physical Properties Using Molecular Connectivity Indices for Application in Modeling the Environmental Fate of Organic Compounds.

Doug Denne, M.S. expected December 1990. Tentative thesis title: Investigating the Relation Between Molecular Structure and Parameters Important in the Environmental Fate of Organic Chemicals.

Mark Holt, M.S. expected December 1991 Tentative thesis title: Microcomputer Program Utilizing Quantitative Structure Activity Relationships for the Estimation of Properties for the Use in Fate and Transport Models.

Rick Miles, M.S. 1992, expected December 1992

REFERENCES

1. Kenaga, E.E. and C.A.I. Goring, Relationship Between Water Solubility, Soil Sorption, Octanol-Water Partitioning, and Concentration of Chemicals in Biota. 1980. American Society for Testing and Materials: p. 78-115.

2. Karickhoff, S.W., Semi-Empirical Estimation of Sorption of Hydrophobic Pollutants on Natural Sediments and Soils. Chemosphere, 1981. 10(8): p. 833-846.

3. Mackay, D., Correlation of Bioconcentration Factors. Environ. Sci. Technol., 1982. 16(5): p. 274-278.

4. Randic, M., On Characterization of Molecular Branching. J. Amer. Chem. Soc., 1972. 97(23): p. 6609-6615.

5. Kier, L.B. and L.H. Hall, Molecular connectivity in chemistry and drug research. in Medicinal Chemistry a Series of Monographs. 1976, N. Y.: DeSteven, G. ed. Academic Press.

6. Kier, L.B., Molecular connectivity as a description of structure for SAR analysis in Physical Chemical Properties of Drugs. 1980, New York, NY: Yalkowsky, S. H., A. A. Sikula, and S. C. Valvani, eds. Marcel Dekker, Inc. 331pp.

7. Kier, L.B., Hall, and L. H., Molecular Connectivity in Structure-Activity Analysis. 1986, Letchworth, Hertfordshire, England.: Research Studies Press, LTD., England, and John Wiley & Sons, Inc., New York. 262.

8. Sabljic, A., Predictions of the Nature and Strength of Soil Sorption of Organic Pollutants by Molecular Topology. J. Agric. Food Chem., 1984. 32: p. 243-246.

9. Sabljic, A., On the Prediction of Soil Sorption Coefficients of Organic Pollutants from Molecular Structure: Application of Molecular Topology Model. Environ. Sci. Technol., 1987. 21: p. 358-366.

10. Doucette, W.J. and A.W. Andren, Estimation of Octanol/Water Partition Coefficients: Evaluation of Six Methods for Highly Hydrophobic Aromatic Hydrocarbons. Chemosphere, 1988. 17(2): p. 334-359.

11. Nirmalakhandan, N.N. and R.E. Speece, QSAR Model for Predicting Henry's Constant. Environ. Sci. Technol., 1988. 22(11): p. 1349-1357.

12. Briggs, G.G., Theoretical and Experimental Relationships between Soil Adsorption, Octanol-Water Partition Coefficients, Water Solubilities, Bioconcentration Factors and the Parachor. J. Agric. Food Chem., 1981. 29: p. 1050-1059.

13. Hermann, R.B., Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with solvent cavity surface area. J. Phys. Chem., 1972. 76: p. 2754-2759.

33

14. Amidon, G.L., S.H. Yalkowsky, and S.Leung, Solubility of Nonelectroytes in Polar Solvents II. Solubility of Aliphatic Alccohols in Water. J. Pharm. Sci., 1974. Vol. 63(No. 12): p. 1858-1866.

15. Yalkowsky, S.H. and S.C. Valvani, Partition Coefficients and Surface Areas of Some Alkylbenzenes. Journal ot Medicinal Chemistry, 1976. Vol. 19, No. 5: p. 727-728.

16. Yalkowsky, S.H. and S.C. Valvani, Solubilities and Partitioning 2. Relationships Between Aqueous Solubilities, Partition Coefficients, and Molecular Surface Areas of Rigid Aromatic Hydrocarbons. Journal of Chem., 1979. 24(2): p. 127-129.

17. Mackay, D., R. Mascarenhas, and W.Y. Shiu, Aqueous Solubility of Polychlorinated Biphenyls. Chemosphere, 1980. 9: p. 257-264.

18. Amidon, G.L. and S.T. Anlk, Application of the Surface Area Approach to the Correlation and Estimation of Aqueous Solubility and Vapor Pressure. Alkyl Aromatic Hydrocarbons. Journal of Chemical and Engineering Data, 1981. Vol. 26 No. 1: p. 28-33.

19. Pearlman, R.S., Molecular Surface Areas and Volumes and Their Use in Structure/Activity Relationships, in Physical Chemical Properites of Drugs, S.H. Yalkowsky, A.A. Sikula, andS.C. Valvani, Editor. 1980, Marcel Dekker, Inc.: New York. p. 331-346.

20. Doucette, W.J. and A.W. Andren, Correlation of Octanol/Water Partition Coefficients and Total Molecular Surface Area of Highly Hydrophobic Aromatic Compounds. Environ. Sci. Technol., 1987. 21: p. 821-824.

21. Hawker, D.W. and D.W. Connell, Octanol-Water Partition Coefficients of Polychlorinated Biphenyl Congeners. Environ. Sci. Technol., 1988. 22(4): p. 382-387.

22. Fugate, H.N., Using Total Molecular Surface Area in Quantitative Structure Activity Relationships to Estimate Environmental Fate and Transport Parameters. 1989, Utah State University:

23. Connolly, M.L., Solvent-Accessible Surfaces of Proteins and Nucleic Acids. Science, 1983. 221(4612): p. 709-713.

24. Fredenslund, A., R.L. Jones, and J.M. Prausnitz, Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures. AIChE Journal, 1975. 21(6): p. 1086-1099.

25. Fredenslund, A.J., Gmehling, and P. Rasmussen, Vapor-liquid equilibria using UNIFAC, a group-contribution method. 1977, New York, NY 213 pp.: Elsevier Scientific Publishing Co.

26. Gmehling, J., P. Rasmussen, and A. Fredenslund, Vapor-liquid equilibria by group contribution. Revision and Extension. 2. Ind. Eng. Chem. Process Des. Dev., 1982. 21: p. 118-127.

27. Arbuckle, W.B., Estimating Activity Coefficients for Use in Calculating Environmental Parameters. Environ. Sci. Technol., 1983. 17: p. 537-542.

28. Banerjee, S., Calculation of Water Solubility of Organic Compounds with UNIFAC-Derived Parameters. Environ. Sci. Technol., 1985. 19(4): p. 369-370.

29. Arbuckle, W.B., Using UNIFAC to Calculate Aqueous Solubilities. Environ. Sci. Technol., 1986. 20: p. 1060-1064.

30. Doucette, W.J., Measurement and Estimation of Octanol/Water Partition Coefficients and Aqueous Solubilities for Halogenated Aromatic Hydrocarbons. 1985, PhD. thesis University of Wisconsin:

31. Banerjee, S. and P.H. Howard, Improved Estimation of solubility and Partitioning Through Correction of UNIFAC-Derived Activity Coefficients. Environ. Sci. Technol., 1988. 22(7): p. 839-841.

32. Mackay, D., Finding Fugacity Feasible. Eviron. Sci. and Technol., 1979. 13(12): p. 1218-1223.

33. Anderson, E., G.D. Veith, and D. Wieninger, SMILES: A Line Notation and Computerized Interpreter for Chemical Structures. 1987. EPA/600/M-87/021.

34. Weininger, D., SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. J. Chem. Inf. Comput. Sci., 1988. 28: p. 31-36.

35. Frazier, J.D., Physical/Chemical Property Estimation Using Molecular Connectivity Indices for Application in Environmental Fate Modeling. 1990, Utah State University:

36. Bahnick, D.A. and W.J. Doucette, Use of Molecular Connectivity Indices to Estimate Soil Sorption Coefficients for Organic Chemicals. Chemosphere, 1988. 17(9): p. 1703-1715.

37. Bondi, A., Van der Waals Volumes and Radii. Journal of Physical Chemistry, 1964. Vol 68(Number 3): p. 441-451.

38. Grain, Activity Coefficients, in Handbook of Chemical and Property Estimation Methods, W.J. Lyman, W.F. Reehl, andD.H. Rosenblatt, Editor. 1982, American Chemical Society: Washington D. C.

39. Mackay, D. and S. Paterson, Calculating Fugacity. Environ. Sci. Technol., 1981. 15(9): p. 1006-1014.